Stat 544 Exam 1

March 13, 2008

I have neither given nor received unauthorized assistance on this examination.

________________________________
signature

________________________________
name printed

There are 12 parts on this exam. I will score every part out of 10 points and take your best 10 of 12 scores. (Budget your time accordingly.)
1. Below are possible distributions for $Y$ depending upon the value of a parameter $\theta \in \{1, 2, 3\}$.

| $y$ | $f(y|1)$ | $y$ | $f(y|2)$ | $y$ | $f(y|3)$ |
|-----|----------|-----|----------|-----|----------|
| 1   | 1/2      | 1   | 1/4      | 1   | 0        |
| 2   | 1/2      | 2   | 1/2      | 2   | 1/2      |
| 3   | 0        | 3   | 1/4      | 3   | 1/2      |

a) Two independent observations $Y_1$ and $Y_2$ (with marginal pmfs $f(y|\theta)$) take values 2 and 3. Based on this an a prior for $\theta$ that is uniform on $\{1, 2, 3\}$, find the posterior distribution for $\theta$.

\[
f(2|\theta)f(3|\theta) = ?
\]
\[
f(2|1)f(3|1) = 0
\]
\[
f(2|2)f(3|2) = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}
\]
\[
f(2|3)f(3|3) = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}
\]

\[
g(\theta|\text{data}) = \begin{cases} 
\frac{1}{3} & \theta=2 \\
\frac{2}{3} & \theta=3 
\end{cases}
\]

b) Based on your answer to a) find a posterior predictive distribution for $Y_{\text{new}}$ (that conditional on $\theta$ is independent of $Y_1$ and $Y_2$ with the same marginal pmf).

This is

\[
\sum_{\theta} f(y_{\text{new}}|\theta) g(\theta|\text{data}) = \frac{1}{3} f(y_{\text{new}}|2) + \frac{2}{3} f(y_{\text{new}}|3)
\]

| $y_{\text{new}}$ | $f(y_{\text{new}}|\text{data})$ |
|-----------------|-------------------------------|
| 1               | $\frac{1}{3} \left( \frac{1}{4} \right) = \frac{1}{12}$ |
| 2               | $\frac{1}{3} \left( \frac{1}{2} \right) + \frac{2}{3} \left( \frac{1}{2} \right) = \frac{1}{2}$ |
| 3               | $\frac{1}{3} \left( \frac{1}{4} \right) + \frac{2}{3} \left( \frac{1}{2} \right) = \frac{5}{12}$ |
c) Bayesians like to argue that as long as their priors "spread the prior probability around sufficiently" (so that they are sure to "cover" any true parameter value) their posteriors will be "consistent" in the sense of piling up around the true value of a parameter with increasing sample information. Suppose that in fact \( \theta = 2 \) and \( Y_1, Y_2, \ldots \) are iid with marginal pmf \( f(\cdot | 2) \). For the uniform prior on \( \{1, 2, 3\} \) here evaluate

\[
P_{\theta = 2} \left[ g(1 | Y_1, Y_2, \ldots, Y_n) > 0 \right] = P_{\theta = 2} \left[ \text{no } Y_i = 3 \right] = (1 - \frac{1}{4})^n \rightarrow 0
\]

\[
P_{\theta = 2} \left[ g(3 | Y_1, Y_2, \ldots, Y_n) > 0 \right] = P_{\theta = 2} \left[ \text{no } Y_i = 1 \right] = (1 - \frac{1}{4})^n \rightarrow 0
\]

(These are \( \theta = 2 \) probabilities that through \( n \) observations, the posterior has not eliminated respectively the possibilities that \( \theta = 1 \) and \( \theta = 3 \).)

d) (Censoring) Suppose that in contrast to the situation in part a), I get to see not the full information in \( (Y_1, Y_2) \), but only the values of

\[
Z_i = \begin{cases} 
0 & \text{if } Y_i = 1 \\
1 & \text{if } Y_i = 2 \text{ or } 3
\end{cases}
\]

for \( i = 1, 2 \). Based still on the uniform prior over \( \{1, 2, 3\} \), but now the information that \( Z_1 = 1 \) and \( Z_2 = 1 \), what is the posterior distribution of \( \theta \)?

\[
P_{\theta = 1} \left[ Z_1 = 1 \text{ and } Z_2 = 1 \right] = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}
\]

\[
P_{\theta = 2} \left[ Z_1 = 1 \text{ and } Z_2 = 1 \right] = \frac{3}{4} \cdot \frac{3}{4} = \frac{9}{16}
\]

\[
P_{\theta = 3} \left[ Z_1 = 1 \text{ and } Z_2 = 1 \right] = 1 \cdot 1 = 1
\]

\[
P_{\theta} \left[ Z_1 = 1 \text{ and } Z_2 = 1 \right] g(\theta) = \frac{1}{3} \times \text{times above}
\]

The sum of these is \( \frac{1}{3} \left( \frac{29}{16} \right) \)

So the posterior is

| \( \theta \) | \( g(\theta | Z_i = 1, Z_2 = 1) \) |
|---|---|
| 1 | \( \frac{4}{29} \) |
| 2 | \( \frac{9}{29} \) |
| 3 | \( \frac{16}{29} \) |
2. Consider the possibility of specifying a "distribution" for \((\theta_1, \theta_2)\) by a "density" on \(R^2\) by
\[
g(\theta_1, \theta_2) \propto \exp\left(- (\theta_1 - \theta_2)^2\right)
\]
a) Argue very carefully that such a \(g(\cdot)\) does NOT specify a proper probability distribution for \((\theta_1, \theta_2)\).

\[
\sqrt{\pi} \cdot \text{normal} \left(\theta_2, \frac{1}{2}\right) \text{ density}
\]
\[
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(- (\theta_1 - \theta_2)^2\right) \, d\theta_1 \, d\theta_2
\]
\[
= \int_{-\infty}^{\infty} \sqrt{\pi} \, d\theta_2 = \infty \quad \text{so} \quad g(\theta_1, \theta_2)
\]
\[
\text{can not be normalized to produce a joint density for } (\theta_1, \theta_2)
\]

b) In spite of the fact in a), someone who wasn't paying attention might attempt to "sample from \(g(\cdot)\)" using a Gibbs sampler. Identify exactly how that person would make updates (from, say, \((\theta_1^i, \theta_2^i)\) to \((\theta_1^{i+1}, \theta_2^{i+1})\)). (What are the two conditionals from which the person would sample?)

Update \(\theta_1^{i+1}\) by sampling from \(h(\theta_1) \propto \exp\left(- (\theta_1 - \theta_2^i)^2\right)\) proportional to a \(N(\theta_2^i, \frac{1}{2})\) density

Similarly, update \(\theta_2^{i+1}\) by sampling from a \(N(\theta_1^{i+1}, \frac{1}{2})\) distribution
c) How do you expect the sampler described in b) to behave (from, say, a start at \( (0,0) \))? Do you expect the (joint) relative frequency distribution of \( \{(\theta_1^i, \theta_2^i)\}_{i=1,...,N} \) to converge? If so, what can you say about the limit? If not, what can you say about what you expect to happen to the relative frequency distribution with increasing \( N \)?

The sequence \( \theta_2^1, \theta_1^2, \theta_2^2, \theta_1^3, \theta_2^3, \theta_1^4, \theta_2^4, \ldots \) is a classical (normal) random walk. This has variance

\[
\text{Var } \theta_2^n = 2(n-1)\left(\frac{1}{2}\right) = n-1 \xrightarrow{n \to \infty} \infty
\]

The empirical means will probably not converge. Both of the marginals will become increasingly spread out as \( n \to \infty \) (as the random walk takes increasingly wide swings above and below its start at \( \theta_2^1 \)).

d) In spite of the fact in a), in a model where conditioned on \( (\theta_1, \theta_2) \) variables \( Y_1 \) and \( Y_2 \) are independent normal variables with means \( \theta_1 \) and \( \theta_2 \) and variance 1, \( g(\cdot) \) can be used as an improper prior and produces a legitimate posterior distribution. Carefully argue that this "works." (Hint: You can bound \( g(\cdot) \) above by some constant.)

\[
\begin{align*}
&f(y_1 | \theta_1) f(y_2 | \theta_2) g(\theta_1, \theta_2) \leq f(y_1 | \theta_1) f(y_2 | \theta_2) \\
&\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y_1-\theta_1)^2\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y_2-\theta_2)^2\right) \, d\theta_1 \, d\theta_2 \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y_2-\theta_2)^2\right) \, d\theta_2 \\
&= 1 < \infty
\end{align*}
\]
3. Suppose that conditioned on positive parameters $\lambda, \delta_1,$ and $\delta_2,$ the variables $Y_1, Y_2,$ and $Y_3$ are independent Poisson variables with respective means $\lambda, \lambda + \delta_1,$ and $\lambda + \delta_1 + \delta_2.$ Suppose that one observes $Y_1 = 7, Y_2 = 5,$ and $Y_3 = 6$ and uses independent $U(0,10)$ priors for $\lambda, \delta_1,$ and $\delta_2.$

a) Describe completely a Metropolis-Hastings or a Metropolis-Hastings-within-Gibbs algorithm that one could use to sample from the posterior distribution of the parameters.

$$L(\delta, \lambda, \lambda_2) = e^{-\lambda} \lambda^{Y_1} (\lambda + \delta_1)^{Y_2} (\lambda + \delta_1 + \delta_2)^{Y_3} / Y_1! Y_2! Y_3!$$

$$\lambda \in \mathbb{R}^+ - 2 \delta_1 - \delta_2 \gtrless \lambda (\lambda + \delta_1)^{Y_2} (\lambda + \delta_1 + \delta_2)^{Y_3}$$

So the posterior is proportional to the above when all of $\lambda, \delta_1, \delta_2$ are in $(0,10)$ and is 0 otherwise.

---

A Metropolis algorithm: Given $(\lambda^*, \delta_1^*, \delta_2^*)$, generate proposal $(\lambda^*, \delta_1^*, \delta_2^*)$ from the prior and then compute the ratio

$$r = \frac{L(\lambda^*, \delta_1^*, \delta_2^*)}{L(\lambda, \delta_1, \delta_2)}$$

and set $(\lambda^*, \delta_1^*, \delta_2^*) = (\lambda^*, \delta_1^*, \delta_2^*)$ with probability $\min (r, 1)$ and set $(\lambda^*, \delta_1^*, \delta_2^*) = (\lambda, \delta_1, \delta_2)$ otherwise.

---

A Metropolis-within-Gibbs algorithm: Given $(\lambda, \delta_1, \delta_2)$

generate $\lambda^* \sim U(0,10)$ and compute the ratio

$$r = \frac{L(\lambda^*, \delta_1, \delta_2)}{L(\lambda, \delta_1, \delta_2)}$$

and set $\lambda^* = \lambda^*$ with probability $\min (r, 1)$ and $\lambda^* = \lambda$ otherwise.

Generate $\delta_1^* \sim U(0,10)$ and compute the ratio

$$r = \frac{L(\lambda, \delta_1^*, \delta_2)}{L(\lambda, \delta_1, \delta_2)}$$

and set $\delta_1^* = \delta_1^*$ with probability $\min (r, 1)$ and $\delta_1^* = \delta_1$ otherwise.

Generate $\delta_2^* \sim U(0,10)$ and compute the ratio

$$r = \frac{L(\lambda, \delta_1, \delta_2^*)}{L(\lambda, \delta_1, \delta_2)}$$

and set $\delta_2^* = \delta_2^*$ with probability $\min (r, 1)$ and $\delta_2^* = \delta_2$ otherwise.
b) Write WinBUGS code for implementing a Gibbs sampler for the posterior distribution of the parameters.

```r
model {

mu1<-lambda
mu2<-lambda+delta1
mu3<-lambda+delta1+delta2

Y1~dpois(mu1)
Y2~dpois(mu2)
Y3~dpois(mu3)

lambda~dunif(0,10)
delta1~dunif(0,10)
delta2~dunif(0,10)
}

list(Y1=7,Y2=5,Y3=6)
```

4. Suppose that someone is interested in the mean of the determinant of a sample covariance matrix for \( n = 20 \) observations from \( \text{MVN}_3(\boldsymbol{\mu}, 2 \boldsymbol{I}) \). Write WinBUGS code that could be used to evaluate this mean. (Degrees of freedom for the sample covariance matrix are 19 and WinBUGS has a log-determinant function `logdet()`.)

```r
model {

W[1:3 , 1:3] ~dwish(Lambda[ , ], 19)

ld<-logdet(W[1:3, 1:3])
det<-exp(ld)/(19*19*19)
}

list(Lambda = structure(.Data = c(.5, 0, 0, 0, .5, 0, 0, .5, .5), .Dim = c(3, 3)))
```

Note here that for \( S \) such a sample covariance matrix,
\[
(n-1)S \sim \text{Wishart}(n-1, 2I)
\]
i.e.
\[
(n-1)S \sim \text{WinBUGS-Wishart}(n-1, 5I)
\]
Further, for a \( 3 \times 3 \) matrix A
\[
\det(cA) = c^3 \det(A)
\]
5. Some angles between holes drilled (using so-called electrical discharge machining) in precision metal parts of a certain type and the flat top surfaces of the parts are supposed to be $45^\circ \pm 2^\circ$. 10 such measured angles are below (measurements to the nearest degree) 

$46, 45, 45, 45, 44, 45, 43, 45, 45, 46$

Supposing that angles are iid $N(\mu, \sigma^2)$, some quantities potentially of interest to the manufacturer are

$$Y_{new} = \frac{47 - \mu}{\sigma} - \Phi \left( \frac{43 - \mu}{\sigma} \right)$$

$$C_p = \frac{47 - 43}{6\sigma}$$

$$C_{pk} = \min \left\{ \frac{47 - \mu}{3\sigma}, \frac{\mu - 43}{3\sigma} \right\} = \frac{(47 - 43) - 2|\mu - 45|}{6\sigma}$$

(an additional angle, the fraction of angles that conform to the requirements, a ratio of spread in requirements to the "spread" in the distribution of angles, and a so-called "process capability ratio").

There are two WinBUGS printouts for analyses of the observed angles following this page. What do these indicate about the extent to which the manufacturer is producing parts with angles meeting the $45^\circ \pm 2^\circ$ requirements?

Both of the analyses indicate there is substantial uncertainty about the fraction of parts with angles outside the engineering requirements of $45^\circ \pm 2^\circ$ but there seems to be a serious possibility that $p(\mu, \sigma^2)$ is not large. Further, it's pretty clear that $C_{pk}$ is not huge ($C_{pk} = 1$ means that the mean diameter is exactly 3 standard deviations inside the closest of the two engineering specifications... here it's clear that $C_{pk}$ is not too much above 1 if it is that large).
model {
mu ~ dflat()
logsigma ~ dflat()
sigma <- exp(logsigma)
tau <- exp(-2*logsigma)
for (i in 1:10) {
  Y[i] ~ dnorm(mu, tau)
}
Ynew ~ dnorm(mu, tau)
prob <- phi((47-mu)/sigma)-phi((43-mu)/sigma)
Cp <- 2/(3*sigma)
Cpk <- ((4-2*abs(mu-45))/(6*sigma))
}
list(Y=c(46,45,45,44,45,43,45,45,46))
list(mu=45, logsigma=0, Ynew=45)

This treats the measurements as "real numbers"
model {
  mu~dflat()
  logsigma~dflat()
  sigma<-exp(logsigma)
  tau<-exp(-2*logsigma)
  X[1]~dnorm(mu,tau) I(45.5,46.5)
  X[2]~dnorm(mu,tau) I(44.5,45.5)
  X[3]~dnorm(mu,tau) I(44.5,45.5)
  X[4]~dnorm(mu,tau) I(44.5,45.5)
  X[5]~dnorm(mu,tau) I(43.5,44.5)
  X[6]~dnorm(mu,tau) I(44.5,45.5)
  X[7]~dnorm(mu,tau) I(42.5,43.5)
  X[8]~dnorm(mu,tau) I(44.5,45.5)
  X[9]~dnorm(mu,tau) I(44.5,45.5)
  X[10]~dnorm(mu,tau) I(45.5,46.5)
  Xnew~dnorm(mu,tau)
  Ynew<-round(Xnew)
  prob<-phi((47-mu)/sigma)-phi((43-mu)/sigma)
  Cp<-2/(3*sigma)
  Cpk<-(4-2*abs(mu-45))/(6*sigma)
}

list(mu=45,logsigma=0,Xnew=45)